Journal of Experimental Child Psychology, accepted 2019

Title: Working memory and high-level cognition in children: An analysis of timing and accuracy in complex span tasks.

Authors:

Rebecca Gordon[1] (1footnote)

James H Smith-Spark[1]

Elizabeth J Newton[1]

Lucy A Henry[2]

[1] London South Bank University
103 Borough Rd, London SE1 0AA, UK
Rebecca Gordon email: Rebecca.gordon@ucl.ac.uk
James H Smith-Spark email: smithspj@lsbu.ac.uk
Elizabeth J Newton email: liz.newton@lsbu.ac.uk

[2]City, University of London
Northampton Square, Clerkenwell, London EC1V 0HB ,UK
Lucy A Henry email: Lucy.Henry.1@city.ac.uk

Corresponding author:
Rebecca Gordon
email: Rebecca.gordon@ucl.ac.uk
UCL Institute of Education, University College London
20 Bedford Way, Bloomsbury, London WC1H 0AL, UK

1

Rebecca Gordon
Current and permanent address:
UCL Institute of Education
University College London
20 Bedford Way
Bloomsbury
London
WC1H 0AL
UK
email: Rebecca.gordon@ucl.ac.uk

Working memory and high-level cognition in children: An analysis of timing and accuracy in complex span tasks.

Abstract

This study examined working memory (WM) using complex span tasks (CSTs) to improve theoretical understanding of the relationship between WM and high-level cognition (HLC) in children. Ninety-two children aged between seven and eight years were tested on three computer-paced CSTs and measures of non-verbal reasoning, reading and mathematics. Processing times in the CSTs were restricted based on individually titrated processing speeds, and performance was compared to participant-led tasks with no time restrictions. Storage, processing accuracy, and both processing and recall times within the CSTs were used as performance indices to understand the effects of time restrictions at a granular level. Restricting processing times did not impair storage, challenging models that argue for a role of maintenance in WM. A task-switching account best explained the effect of time restrictions on performance indices and their inter-relationships. Principal component analysis showed that a single factor with all performance indices from just one CST (Counting span) was the best predictor of HLC. Storage in both the participant-led and computer-paced versions of this task explained unique and shared variance in HLC. However, the latter accounted for more variance in HLC when contributions from processing time were included in the model. Processing time in this condition also explained variance above and beyond storage. This suggests that faster processing is important to keep information active in WM; however, this is only evident when time restrictions are placed on the task and important when WM performance is applied in broader contexts that rely on this resource.

Key words: Working memory; High-level cognition; Processing speed

Introduction

Working memory (WM) is commonly defined as the cognitive system responsible for the temporary storage and processing of information. Understanding individual differences in children's WM in primary school is important because they can explain variability in high-level cognition (HLC) including reading (Gathercole et al., 2006; Seigneuric et al., 2000; Towse et al., 2008) and mathematics (Alloway & Passolunghi, 2011; Berg, 2008; Bull & Scerif, 2001; Cragg et al., 2017; Swanson & Beebe-Frankenberger, 2004). Similarly, WM deficits in primary school children are linked to mathematical learning difficulties (Andersson & Lyxell, 2007; Geary, et al., 2007; Iuculano et al., 2011; Passolunghi & Cornoldi, 2008; Passolunghi & Siegel, 2004), reading disabilities (Gathercole et al., 2006), language impairments (Henry & Botting, 2017), and general learning difficulties (Gathercole & Pickering, 2000; Henry & MacLean, 2002).

Empirical investigation of models of WM can address controversies regarding the value and effectiveness of metacognitive WM strategies training to improve classroom performance (e.g. Partanen et al., 2015; Gathercole et al., 2012; Holmes & Gathercole, 2014; Shipstead et al., 2012). Although there is agreement that WM is responsible for the coordination of processing and storage, there are different accounts of how this system operates (see Gathercole & Alloway, 2006, for a review). The influential multi-component model of WM (Baddeley, 1986; 2000; Baddeley & Hitch, 1974) describes a modality-free control system (i.e. the central executive) with two modality-specific sub-systems responsible for the temporary storage of phonological and visuospatial material. Processing and storage share resources from the central executive; however, as the resources are limited, increased memory load reduces capacity for processing and, conversely, increased cognitive load during processing (e.g. more complex or numerous stimuli) reduces capacity for storage. However, according to the model, storage capacity can be boosted when these two sub-systems actively maintain memoranda via verbal rehearsal of phonological information (Baddeley, 1986) and image generation for visuospatial information (Logie, 2014).

Based on this model, studies have examined WM capacity using complex span tasks (CSTs) designed to replicate the requirement to temporarily maintain and manipulate information. For instance, Counting Span (Case et al., 1982) requires participants to process information (counting shapes) and store memoranda (number of shapes presented). The number of items to be stored increases across trials and the total number correctly recalled yields a span score which reliably reflects WM capacity (Conway et al., 2005).

Using this task, Case et al. (1982) found that children's storage capacity was a function of the speed with which they could count the array of objects. They argued that more efficient processing frees up cognitive resources for storage resulting in higher span scores. This explanation of WM is referred to as the resource-sharing hypothesis. However, this account was challenged by Towse and Hitch (1995; see also Towse et al. 1998) who manipulated both processing complexity and time. Consistent with Case et al. (1982), it was found that higher span scores related to faster processing (i.e. counting). However, increasing the difficulty of the processing component did not reduce storage. It was argued that children switch away from storage during the processing, as opposed to sharing a single cognitive resource to undertake both processes. This task-switching account posits that storage in WM is not predominantly determined by resources taken up by processing, but by time-based forgetting as a function of length of time spent on processing.

Cowan and colleagues (Cowan, 1999, 2008; Cowan et al., 2005; 2015) propose an alternative to the Baddeley & Hitch (1974) account of WM with the embedded-process model. The premise is that WM uses attentional resources to activate information from a single, central memory store (Cowan, 1995; Cowan et al., 2010). Furthermore, although the embedded-process model notes the importance of processing (i.e. attention) and storage (i.e. activation) (Cowan et al., 1999), an interplay between these constructs is not emphasised. Rather, Cowan et al. (2005) see the role of attention as fundamental to WM capacity. To demonstrate this, they compared performance in children and adults on a memory task that required both processing of information and memory for spatial locations. Participants were presented with items to attend to (hits) and items to ignore (false alarms). It was found that when the arrays were small, adults and children were equally able to favour recall of hits rather than alarms, denoting a comparable attentional efficiency (i.e. processing). However, the total number of items remembered (i.e. hits and false alarms) was lower for children than adults, and when the size of the arrays increased, only the children's attentional efficiency was impaired (i.e. more false alarms were recalled). Based on these findings, Cowan and colleagues posit that a core attentional capacity explains differences in WM.

Barrouillet et al. (2004) further investigated the importance of attention in WM, demonstrating that diverting attention away from active maintenance has a detrimental effect on the recall of memoranda. This, the Time-Based Resource-Sharing (TBRS) model describes a limited attentional resource that switches between storage and processing of information to keep information active in WM. By manipulating cognitive load to increase processing time in CSTs, Barrouillet and colleagues (Barrouillet et al., 2009; 2011) demonstrated that the time taken to process stimuli leads to memory decay and that this is more important than time allowed for active maintenance of memoranda. This was demonstrated in a negative linear relationship between processing time and storage scores. Despite some evidence of the optional use of rehearsal (Camos et al., 2011), the TBRS model argues for the importance of opportunities for attentional refreshing of storage items in WM. This was evident when increasing the pace of delivery of processing stimuli in CSTs had a deleterious effect on children's recall (Camos & Barrouillet, 2011; Lépine et al., 2005). This, they argued, was because a faster pace reduces opportunity to refresh memoranda during small gaps between processing items.

Thus far, the following explanations for differences in WM have been presented: active maintenance (Baddeley, 1986; Logie, 1995), resource-sharing (Case et al., 1982), task-switching (Towse & Hitch, 1995), core attentional capacity (Cowan et al., 2005), and attentional refreshing (Barrouillet et al., 2004). Research with adults has examined WM to improve theoretical understanding of the different WM models and subsequent relationships with HLC. There are two key approaches that are relevant to the current study: 1) controlling processing time within CSTs and 2) examining CST performance indices beyond storage.

With regard to the first approach, computer-paced tasks have been used to restrict processing times within CSTs and found that partialling out variance explained in HLC in participant-led and computer-paced conditions has shown these respective tasks measure both similar and different abilities (Bailey, 2012, Unsworth et al., 2005). By analysing processing times within the tasks, Unsworth et al. (2005) found that the computer-paced task explained variance in HLC above and beyond storage, whereas this was not the case in the participant-led task. Similarly, Friedman & Miyake (2004) found that processing times in participant-led and experimenter-led reading span tasks correlated with span scores; however, the longer processing times in the participant-led task weakened correlations with reading ability compared to the experimenter-led task (see also St Clair-Thompson 2007). These findings support the embedded-process model, which argues that time for maintenance in WM tasks introduces individual variation in cognitive abilities unimportant in the WM-HLC relationship (Cowan et al., 1999).

This is in line with findings from a study with children that manipulated CST processing times whilst controlling for individual differences in processing speed to further understand the WM-HLC relationship. Lépine et al. (2005) compared performance by 11-year-olds on automated and participant-led CSTs in which the presentation length of the processing stimuli was either based on a generic time limit (e.g. 1,000ms) or items were presented for as long as it took for the participant to process the stimuli. In line with adult results, the time-restricted tasks showed significantly stronger links to HLC compared to the participant-led tasks. The authors argued that time-restricted tasks provide a purer measure of WM, less influenced by other cognitive abilities invoked by maintenance, and that this fundamental capacity predicts HLC (see also, Cowan et al., 1999). However, Lépine et al. employed the *same* time duration for automated presentation for all participants, not accounting for individual differences in processing speed. Processing stimuli more quickly than this may have freed up time for maintenance of memoranda before the next step of the task. Conversely, participants who processed stimuli more slowly would not have been able to perform the processing task and would therefore fail. Therefore, the same degree of constraint was not applied to all participants.

Another approach previously used to investigate the WM-HLC relationship relevant to this study is the examination of CST performance indices beyond storage. Previous studies with children have demonstrated that recall time (Towse et al., 2008) and processing times (Bayliss et al., 2003; 2005) within CSTs predict HLC. However, compared to research with adults (e.g. Unsworth et al., 2005), there is little research with children that has unpacked CSTs to better understand the mechanisms of WM, how they explain individual differences in capacity, and why they predict HLC so well. Given that WM is a good predictor of HLC in children (e.g. Cragg et al., 2017, Henry & Botting, 2017), there is a need to provide further explanation of individual differences in WM. In doing so, further insight into the aforementioned models of WM can be provided. This may inform intervention strategies that aim to boost academic achievement in children (e.g. Ribner et al., 2017).

The current study

The current study unpacked how different methodological approaches affect all component performance indices within CSTs (i.e. storage, processing time, recall time, processing accuracy) in order to further our understanding of their relationships with HLC. This was achieved in two ways: 1) by controlling for individual differences in processing speed within CSTs and 2) by unpacking CSTs and examining performance indices beyond storage.

Two major limitations of previous research into CST-HLC relationships were addressed. First, previous research using time-restrictions has not accounted for individual differences in processing speeds, so it is not known whether such differences affect CST performance. Second, previous research has not considered all CST indices: storage, processing time, processing accuracy, and recall time. Therefore, it is unknown whether CST components beyond the typical measure of storage explain variance in HLC.

To measure WM, Counting, Listening and Odd-one-out span tasks were administered in computer-paced conditions where processing times were titrated based on individual processing speeds; and participant-led conditions where there was no such restriction. This method also permitted the extraction of accuracy and speed measures related to processing and storage. The contribution of latent factors to variance in measures of HLC was examined and compared across the two administration conditions.

There are many approaches to identifying active maintenance in WM in young children, including video analysis to detect sub-vocal rehearsal (Lehmann & Hasselhorn, 2007), strategy training (Miller et al., 2015), and manipulation of verbal and non-verbal stimuli (Henry, 1991; 2008). Restricting time allowance for processing in CSTs has been used effectively to identify maintenance use in adults (Bailey, 2012; Friedman & Miyake, 2004; St Clair-Thompson, 2007) and children (Camos & Barrouillet, 2011; Lépine et al., 2005). The aim of the current study was to use this manipulation to explain variation in WM between the two task conditions.

The age-group (seven-year-olds) for this study is of particular importance because research has shown that verbal rehearsal emerges at approximately this point (Gathercole & Hitch, 1993; Gathercole et al., 1994; Henry & Millar, 1991; 1993; but see Jarrold & Citröen, 2012). Similarly, the TBRS model of WM argues for the emergence of an attentional switching capability that explains increased WM capacity at approximately seven years (Camos & Barrouillet, 2011). Thus, this age-group provides an appropriate window to investigate whether controlling for individual differences in processing times affects WM and, by unpacking performance indices within tasks, whether it is possible to identify the source of HLC relationships. Two research questions were explored:

1. What is the effect of controlling for individual differences in processing speeds on CST performance in seven-year-old children compared to performance on tasks with no such restriction?
2. How do measures beyond storage in CSTs play a role in explaining individual differences in children's HLC in these two conditions?

Based on theories of WM discussed thus far, the following suppositions were made. If active maintenance is important in WM (Baddeley, 1986; Baddeley & Hitch, 1974; Logie, 1995) then time-restricted tasks that limit opportunities for maintenance should result in reduced storage and weaken HLC links.

Similarly, if small gaps during processing allow for attentional switching to refresh memoranda (Barrouillet et al., 2004) then time-restricted tasks should reduce opportunities to refresh memory items, resulting in lower storage scores and weaker links with HLC (Camos & Barrouillet, 2011). This should be evident in a significant negative relationship between storage and processing times only in the participant-led condition (Barrouillet et al., 2007). In addition, should processing speed be important in downstream tasks that rely on WM, processing times should be related to HLC.

Conversely, if resource-sharing (Case et al., 1982) or task-switching (Towse & Hitch, 1995; Towse, Hitch & Hutton, 1998) explain WM capacity, then the restriction of processing times should not affect storage, as the constraints were based on individual processing speeds allowing each child a comfortable amount of time to carry out the processing, but not more than they need. Thus, it could be assumed that the children are not restricted per se, but provided with the required amount of time to complete the task. Additionally, no significant difference in the relationship with HLC should be observed between the two conditions. However, a negative relationship between processing times and storage would be expected in both conditions, demonstrating either resource-sharing or task-switching. Furthermore, if faster processing explains greater storage capacity, and this in turn relates to HLC, then processing times should predict HLC in both conditions.

If a core attentional capacity underpins WM (Cowan et al., 2005), then time-restricted tasks that reduce maintenance opportunity should produce cleaner measures of WM and thereby strengthen links with HLC compared with participant-led tasks.

Although there were no specific predictions related to recall time and processing accuracy, they were included as performance indices to ensure a full picture of the component processes involved in CSTs was achieved.

Method

Design

This correlational study explored relationships between CSTs, in two administration conditions (computer-paced, participant-led), and measures of HLC (non-verbal reasoning, reading and mathematics). The data were further analysed using principal component analysis and hierarchical regression to determine the amount of variance accounted for in HLC by indices (storage, processing time, processing accuracy, and recall time) within the participant-led and computer-paced CSTs.

Participants

Ninety-nine participants from Grade 3 primary school were recruited from two South-East London schools. To assess a representative sample of children in UK mainstream education, those with known developmental delays and/or Special Educational Needs statements were excluded. Six children transferred to another school before completing testing. One further child was excluded when identified as colour-blind and unable to complete the Raven’s Coloured-Progressive Matrices. The remaining 92 children (41 male; mean age = 7 years, 10 months, *SD* = 4.23) participated in all five testing sessions; all were unfamiliar with the assessments prior to the commencement of testing.

Ethical Approval and Consent

The study was approved by the University Research Ethics Committee at the authors’ host institution. Written consent was obtained from schools and parents for all participants. Digitally recorded verbal assent to participate was obtained from each child prior to commencement of the first testing session.

Materials

Counting, Listening and Odd-one-out span were administered in participant-led and computer-paced conditions. Both versions were computerised to ensure comparable testing environments. All tasks were presented, either aurally or visually, via a Dell 5000 Series Inspiron laptop, and were written in E-Prime Version 2.0 (Schneider, Eschman, & Zuccolotto 2002). Each task was driven by a push-button response box operated by the researcher.

Counting span was based on Counting Recall from the Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001). The processing component of the task required participants to count an array of either four, five, six or seven dots on the computer screen and say the number out loud to the researcher who recorded the response by pressing the corresponding button on the box. After a block of six trials, the number of stimuli increased to two screens of dots to-be-counted. At the end of each trial the participant was asked to recall how many dots had been on each screen in serial order. The number of screens increased every six trials, up to a maximum of seven screens, or until the participant failed to recall more than three trials out of a block of six. Total trials correct (out of a maximum of 42) represented the participants' storage score on this task.

Listening span was based on Listening Recall from the WMTB-C (Pickering & Gathercole, 2001). For the processing component, participants listened to a sentence (e.g. "Apples have noses"), decided whether it made sense and informed the researcher of their decision by saying "yes" or, as here, "no''. There were 42 sentence stimuli, four to six syllables/words in length. Of these, 50% were nonsensical and the others were true (e.g. ''You sleep in a bed''). The duration of each spoken sentence was two seconds. The sentences were taken from the WMTB-C and an adaptation by Leather and Henry (1994). The researcher recorded the response by pressing the corresponding button on the box. At the end of each block of trials the participant was required to recall the last word of each sentence (in the previous example, "bed") in correct serial order. The experimenter recorded these responses on paper and pressed a button on the box to record the time of response. The number of sentences increased in subsequent blocks as per the Counting span task. The same scoring protocol was used.

Odd-one-out span was based on a task created by Henry (2001) to measure non-verbal WM. The processing component required participants to identify the sole incongruent shape from a horizontal line of three shapes in three separate boxes, The odd one out was always easily identifiable without being immediately obvious (e.g. two arrows pointing to the left and one pointing to the right). The recall component required pointing out the spatial location of the odd-one-out within empty boxes after it had disappeared. The spatial position of the odd-one-out varied across trials, with repetition of the same location within a block on some occasions. The participants were told not to indicate the location verbally to maintain the visuospatial nature of the task. The blocks increased incrementally as per the other CSTs, with the same scoring protocol.

Reading ability (single word decoding) was measured using the Word Reading task from The British Ability Scales Third Edition (BAS-3, Elliot & Smith, 2011). Raw scores were converted to ability scores and then to standardised measures to provide an overall Word Reading score.

Due to differences in curricula across schools, the use of standardised measures of mathematics ability (e.g. Access, BAS-3 Number Skills) would have led to performance differences attributable to variations in exposure to certain topics, not just individual differences in ability. The UK Standard Assessment Tasks (SATs; Kirkup, Sizmur, Sturman & Lewis, 2005) scores for mathematics ability were therefore used, since these provide an assessment of ability relative to learning opportunities. These are based on a framework for teaching mathematics dictated by the UK Government’s Department for Education (DfES, 2003) and are designed to consider the taught topics for that academic year (for a similar approach see Gathercole & Pickering, 2000; Lépine et al., 2005; St Clair-Thompson & Gathercole, 2006). Grades were transformed into single numbers representing each level of ability that was assigned as a SAT score (1 = low through to 12 = high ability).

Non-verbal reasoning ability was measured using the Raven's Coloured Progressive Matrices (RCPM; Raven, 2008). Raw scores were used to obtain a standardised overall score.

Procedure

For each CST, the participants were first required to complete a series of 20 non-memory trials in order to calculate individual processing speeds for the computer-paced conditions. Although this procedure was not necessary for the participant-led condition, it was included to ensure consistency of administration experience. Participants were requested to complete these trials as quickly and carefully as possible. Using counting span as an example, they were presented with a screen displaying an array of dots to be counted out loud, telling the researcher the sum of the count verbally. When participants articulated the final count, the researcher pressed the corresponding button on the box to record the processing time. Timing began from when the screen first appeared and ended when the response was given To avoid carry-over effects, for the Listening span and Odd-one-out span trials, stimuli were used that would not be included in the CST. This was not possible for Counting span due to the limited stimuli pool available.

After the non-memory trials, the program calculated each participant's mean processing time based on their time taken to engage in the processing tasks and provide a response. A minimum of 85% accuracy was required for inclusion in further assessment. This cut-off was based on the automated OSPAN task developed by Unsworth et al. (2005) and was designed to ensure participants were attending sufficiently to the stimuli. In the current study, no participant performed below this level.

For the computer-paced versions, mean processing time plus 2.5 *SD* was used as a time limit for the processing component of the WM tasks (e.g. counting dots). This formula was again based on the automated OSPAN task (Unsworth et al., 2005) and was designed to provide participants with a response window equal to approximately 98% of their individual response times in the non-memory trials. To allow for the variation in speed caused by different quantities of dots on each screen in the Counting span task, a mean duration was calculated for each of the four different counting screens (i.e. four, five, six or seven dots; five screens for each of the four quantities) presented in the non-memory trials.

For each WM task, after the 20 non-memory trials, a practice session was conducted. For participant-led trials, the processing stimulus was presented until a response was made. For computer-paced trials, the processing stimulus was presented for the duration of the individual's time limit. During a 750ms delay a fixation point was displayed on the screen before the next processing stimulus was presented. If the allotted time was exceeded on computer-paced trials, the task moved on to the next step (either the next processing item or the recall stage) and that trial was counted as an error. For the Counting span task, the time limit for each quantity count was applied to the corresponding array.

For all these trials, participants performed the processing component and were then asked to recall the output (e.g. the number of dots) at their own pace at the end of each trial. There were two practice trials, starting with list lengths of one item then increasing to two items. Participants were required to complete all practice trials correctly before moving on to the measurement task. No child failed to complete this step.

All the WM tasks were conducted in the same manner in both conditions, with one exception. For the computer-paced condition, the participants were informed of the time restriction. For example, in the Counting span task they were told: “When you see the screen of dots that you need to count, I want you to start counting them straight away as you only have enough time to count them. If you don’t count them straight away, the computer may move on to the next screen before you have finished”. For the participant-led condition, participants were told: “When you see the screen of dots I want you to count them and tell me how many there are” (or ‘’tell me if the sentence makes sense” or “point to the odd-one-out” for Listening span and Odd-one-out span respectively). As processing time allowance was based on individual processing speeds, faster participants did not have unfilled intervals between completing processing and starting the next trial (or recall). This reduces the possibility they were afforded more maintenance time compared with participants who processed the stimuli more slowly.

With the exception of the SATs mathematics grades, which were collected from the class teachers at the end of Grade 3, the remaining eight tasks were administered throughout the same academic year by the first author. The mean duration between the participant-led version of each task (i.e. session 1) and the computer-paced version (i.e. session 2) was 6.74 (*SD* = 3.60) weeks. For each participant the tasks were presented in the same order, single tasks were always completed in one session, and the entire session was always completed within two school days. The sequence of task administration is shown in Table 1. In each school, testing took place in a quiet area away from distractions of other children and teaching activity.

Table 1. Sequence of tasks within each testing session.

Session	Tasks
1	1. BAS-3 Reading
	2. Counting span (participant-led)
	3. Odd-one-out span (participant-led)
	4. Listening span (participant-led)
2	1. Counting span (computer-paced)
	2. Odd-one-out span (computer-paced)
	3. Listening span (computer-paced)
3	1. Raven's Progressive Colour Matrices

Calculation of WM performance indices

Recall time for each trial was calculated from the time the recall prompt appeared on the screen to recording of the final recall response on the button box. For each block, a composite was calculated from all six trials. Recall time was participant-led regardless of administration condition. Processing accuracy was calculated as the total number of possible correct processing responses minus the total number of errors. Processing time was calculated by summing the total time taken to process each stimulus within a trial, then the mean processing time across trials was calculated for each block.

Due to individual differences in span, not all participants progressed equally far through the seven blocks of trials in the CSTs. Therefore, for recall time, processing time, and processing accuracy, some participants only produced data for the first three blocks before they failed the task. To ensure that all cases were included in the analysis, only data from blocks 1, 2, and 3 were included to create a composite measure for processing time, recall time and processing accuracy for each CST. To remove the influence of any extreme responses (Ratcliff, 1993), the values for recall time and processing time were converted to z-scores to identify any values more than 2.5 SDs from the mean. The corresponding raw values more than

2.5 SDs from the mean for each individual item were winsorized and substituted with the outermost criterion value for that item. This resulted in the alteration of three values of recall time scores across Counting, Listening and Odd-one-out span in the participant-led condition and one in the Listening span computer-paced condition. This totalled 4.3% of data across the sample (for a similar methodology see Bayliss et al., 2003; 2005).

The performance index for storage was total trials correct (TTC) across all blocks to ensure that maximum storage ability was reflected in the analysis. This was consistent for all tasks in each administration condition.

Results

The results are presented in four sections. The first comprises descriptive statistics, missing data report, and reliability analysis for CST performance indices. In the second, the results of t-tests to assess the effect of the time restrictions on the CSTs are presented. The third section considers the results of the principal component analysis (PCA) used to identify CST factors. The results of the regression analyses regarding relationships with HLC are addressed in section four.

1. Descriptive statistics

Means and SDs for storage scores, recall time, processing time, processing accuracy, non-verbal reasoning, reading and mathematics are shown in Table 2. The table also includes an indication of data missing due to procedural error and occasional equipment failure. With regard to the latter, 24 storage scores for the participant-led version of Odd-one-out span and 14 storage scores for the computer-paced version failed to record. The resultant sample size for these two tasks was 68 and 78 respectively. Therefore, the strength of the analyses using these data was weaker compared with storage scores from both versions of the Counting and

Listening span tasks (*n* = 90-92 per task). However, further analysis demonstrated that reliability was robust for the participant-led (α = .71) and computer-paced (α = .73) Odd-one-out span tasks. For all missing values, Little's MCAR test indicated that the missing data could be considered random (χ^2 (15) = 22.329, p = .099). Also, there were random individuals missing data points for participants in the Counting and Listening span tasks. All missing values are reflected in the degrees of freedom for the relevant analyses.

Table 2. Mean scores and standard deviations (SD) for storage, recall time, processing time, and processing accuracy for each CST.

	Complex span tasks					
	Counting span *(SD)*		**Listening span *(SD)***		**Odd-one-out span *(SD)***	
	PL	CP	PL	CP	PL	CP
Storage (TTC)	21.98 *(4.95)* **2*	22.80 *(4.87)* **1*	10.43 *(2.80)* **0*	13.30 *(3.14)* **1*	13.43 *(3.12)* **24*	13.30 *(2.30)* **14*
Recall time (ms)	1297.10 *(412.82)* *2	1063.63 *(373.133)* *1	10133.57 *(2942.75)* *2	7401.36 *(2454.19)* *1	3597.85 *(891.60)* *0	2773.18 *(667.50)* *1
Processing time (ms)	2832.57 *(789.88)* *2	1856.52 *(545.47)* *1	5062.25 *(585.72)* *2	4486.08 *(409.22)* **4*	2800.26 *(480.10)* *0	1975.05 *(361.84)* *1
Processing accuracy (pc)	.99 *(.02)* *2	.89 *(.10)* *1	.96 *(.04)* *2	.94 *(.04)* *4	.98 *(.04)* *0	.94 *(.06)* *1

High-level cognition		
NVR *(SD)*	**Reading *(SD)***	**Mathematics *(SD)***
111.20 *(16.09)* *0	110.66 *(9.70)* *0	8.27 *(1.37)* *0

PL = Participant-led; CP= computer-paced; TTC = total trials correct; ms = milliseconds; pc = proportion correct, NVR = Non-verbal reasoning; *missing number of cases

As only the first three blocks in each span task were used to calculate processing time, recall time and processing accuracy performance indices, significant variations in a score could indicate inconsistencies in the calculation. Therefore, a series of 3 x 2 x 3 dependent analyses of variance (ANOVAs) were conducted to examine the effect of task (Counting, Listening, Odd-one-out), condition (participant-led and computer-paced) and block (1, 2, 3) on processing times, recall times and processing accuracy. Due to the number of variables (six tasks, two conditions, three blocks), the α-level for significance was set at $p < .01$ (for a similar methodology see Geary et al., 2007). The results of each of the three ANOVAs is shown in Table 3. None of the findings were significant, demonstrating there was no systematic variation across task, condition or block at the level of individual blocks for any of the three indices. In addition, there were no significant interactions between any of the three factors. Based on these analyses, the calculation of these performance indices across the three tasks in each administration condition was deemed consistent, and they were used to reflect processing time, recall time and processing accuracy.

Table 3. Analysis of variance for task, condition and block for each performance index

	Processing time	*Recall time*	*Processing accuracy*
Task	*F* (2,100) = .160, *p* = .852, η_p^2 = .003	*F* (2,104) = .109, *p* = .897, η_p^2 = .002	*F* (2,24) = .298, *p* = .745, η_p^2 = .024
Condition	*F* (1,50) = .012, *p* = .913, ηp^2 = .001	*F* (1,52) = .096, *p* = .758, ηp^2 = .002	*F* (1,12) = .143, *p* = .712, ηp^2 = .012
Block	*F* (2,100) = .011, *p* = .989, η_p^2 = .001	*F* (2,104) = 3.290, *p* = .041, η_p^2 = .060	*F* (2,24) = 1.321, *p* = .286, η_p^2 = .099
Task x condition	*F* (2,100) = .343, *p* = .711, η_p^2 = .007	*F* (2,104) = .488, *p* = .616, η_p^2 = .009	*F* (2,24) = 1.053, *p* = .364, η_p^2 = .081
Task x block	*F* (4,200) = 1.226, *p* = .301, η_p^2 = .024	*F* (4,208) = 1.229, *p* = .300, η_p^2 = .023	*F* (4,48) = 1.740, *p* = .157, η_p^2 = .127
Condition x block	*F* (2,100) = 1.366, *p* = .260, η_p^2 = .027	*F* (2,104) = .607, *p* = .547, η_p^2 = .012	*F* (2,24) = .358, *p* = .702, η_p^2 = .029
Task x condition x block	*F* (4,200) = .196, *p* = .940, η_p^2 = .004	*F* (4,208) = 1.082, *p* = .367, η_p^2 = .020	*F* (4,48) = .912, *p* = .465, η_p^2 = .071

To assess the reliability of the storage measure for each CST, a trial-based span score was calculated for all participants. Correct recall on all the first trials was considered (i.e. trial 1 in Block 1, trial 1 in Block 2, trial 1 in Block 3 etc., up to Block 7) until the first trial within a block was not correctly recalled. This was repeated for all trial 2s, trial 3s, etc. up to trial 6. For example, if a participant recalled the first trials in Block 1, Block 2 and Block 3, but not in Block 4, they were awarded a score of '3' (i.e. Block 1 (trial 1) + Block 2 (trial 1) + Block 3 (trial 1) = 3). A score was allocated based on the sum of all correctly recalled trials (i.e. all first trials across all completed blocks, all second trials across all completed blocks etc.). This total was used to denote a span score for each trial. In addition, TTC scores for each span measure were included. Correlational analyses were conducted on all these scores (i.e. all trial spans and TTC) to estimate reliability (for similar methodology see Henry & MacLean, 2002; Engle, Tuholski, Laughlin & Conway, 1999). The correlations between each of the measures all indicated moderate to good task reliability (α = .65 to .78). As a further measure of reliability, TTC for each of the six WM tasks were subjected to split-half reliability analysis. Cronbach's alpha across all tasks showed high reliability (α = .80). Test-retest analyses between the two versions of the counting (α = .72), listening (α = .69) and Odd-one-out (α = .69) span tasks also indicated adequate reliability.

2. Effect of time restrictions on the CSTs

Paired samples *t*-tests compared performance in the two administration conditions (participant-led, computer-paced) to assess whether imposed time restrictions affected overall storage, processing time, recall time and processing accuracy. The results are shown in Table 4. Processing time and recall time were both significantly faster in the computer-paced condition for all three CSTs. Processing accuracy was significantly lower in the computer-paced condition compared to the participant-led condition for all three CSTs. Time restrictions did not result in reduced storage scores for Counting span and Odd-one-out span, but there were significantly higher storage scores in the computer-paced condition of the Listening span task. This unexpected finding is addressed in the discussion.

Table 4. t-test (df) statistics comparing mean scores for storage (total trials correct), processing time (ms), recall time (ms) and processing accuracy (proportion correct) between condition in each CST.

	Storage ***(df) t***	Processing time ***(df) t***	Recall time ***(df) t***	Processing accuracy ***(df) t***
Counting	*(89)* 1.61	*(89)* 15.22	*(89)* 5.40	*(89)* 9.63
	$P = .112$	$P < .001$	$P < .001$	$P < .001$
Listening	*(90)* 10.43	*(85)* 10.24	*(88)* 8.06	*(85)* 3.26
	$P < .001$	$P < .001$	$P < .001$	$P < .01$
Odd-one-out	*(60) 0*.31	*(90)* 15.67	*(90)* 8.40	*(90)* 6.28
	$P = .755$	$P < .001$	$P < .001$	$P < .001$

In order to understand the effect of time restrictions on performance within the CSTs, it was important to consider whether individual performance indices related to each other differently in the two conditions. Table 5 shows that, with the exception of links between processing accuracy and recall time, all relationships between indices were significant for Counting span in both conditions. However, this was not the case for the other tasks. For Listening span the only significant relationships were in the computer-paced condition (storage with recall time, storage with processing accuracy, processing time with recall time). For Odd-one-out span, processing time related to storage and recall time in both conditions. Recall time was linked to storage and processing accuracy in the computer-paced condition only. Processing accuracy was related to storage and processing time only in the participant-led condition.

Table 5. Correlations (df) between performance indices within each complex span task

	Storage	Processing time	Recall time	Processing accuracy
Counting span				
Storage	-	*(89)* -.500**	*(89)* -.360**	*(89)* .314**
Processing time	*(88)* -.645**	-	*(89)* .389**	*(89)* .209*
Recall time	*(88)* -.452**	*(88)* .566**	-	*(89)* -.080
Processing accuracy	*(88)* .399**	*(88)* -.266*	*(88)* -.055	-
Listening span				
Storage	-	*(86)* -0.183	*(89)* -.335**	*(86)* .216*
Processing time	*(88)* -.135	-	*(86)* .300**	*(86)* -.124
Recall time	*(88)* .125	*(88)* .149	-	*(86)* -.075
Processing accuracy	*(88)* .168	*(88)* .096	*(88)* -.024	-
Odd-one-out span				
Storage	-	*(76)* -.301**	*(76)* -.274*	*(76)* .019
Processing time	*(66)* -.408**	-	*(89)* .323**	*(89)* .014
Recall time	*(66)* -.207	*(90)* .530**	-	*(89)* -.305**
Processing accuracy	*(66)* .401**	*(90)* -.263*	*(90)* .030	-

Participant-led below the diagonal; Computer-paced above the diagonal; * = $p < .01$; ** = $p < .001$; *** = $p < .0001$

3. Principal component analysis

The data were analysed to ascertain whether all four performance indices could be identified as separate factors. To establish initial suitability for PCA, correlation analyses were conducted to understand the relationship between the performance indices in each administration condition and task (e.g. storage from Counting span, Listening span and Odd-one-out span in the participant-led condition and then in the computer-paced condition). As recommended by Field (2017), low but significant correlations are required for PCA. Table 6 illustrates that in all but two cases the storage, processing time and recall time in the participant-led and computer-paced tasks were moderately related to each other for all CSTs. However, for processing accuracy there was only one weak correlation.

Table 6. Correlations (df) between performance indices across task and condition

	Storage			Processing time		
	Counting span	**Listening span**	**Odd-one-out span**	**Counting span**	**Listening span**	**Odd-one-out span**
Counting span	-	*(89)* .310**	*(76)* .435**	-	*(86)* .084	*(89)* .360**
Listening span	*(88)* .331**	-	*(76)* .399**	*(86)* .252*	-	*(88)* .444**
Odd-one-out span	*(65)* .272*	*(66)* .211	-	*(90)* .329**	*(90)* .475**	-
	Recall time			**Processing accuracy**		
	Counting span	**Listening span**	**Odd-one-out span**	**Counting span**	**Listening span**	**Odd-one-out span**
Counting span	-	*(89)* .342**	*(89)* .284**	-	*(86)* -.068	*(89)* .191
Listening span	*(86)* .320**	-	*(89)* .299**	*(86)* .252*	-	*(86)* .106
Odd-one-out span	*(88)* .247*	*(88)* .211*	-	*(88)* .196	*(88)* .180	-

Participant-led below the diagonal; Computer-paced above the diagonal; * = p < .01; ** = p < .001; *** = p < .0001

Reliability analyses were conducted separately for each performance index (storage, processing time, recall time, processing accuracy) to determine an adequate association between the three tasks (Counting, Listening and Odd-one-out span) across administration conditions. Cronbach's alpha for recall time (α = .55) and processing accuracy (α = .30) were considered too low to be used for factor analysis so were excluded from further analysis (Tolmie, Muijs & McAteer, 2011). Cronbach's alpha for processing time (α = .70) and storage (α = .77) showed adequate reliability and were therefore included in further analysis to establish factors representing these two performance indices.

PCA was used to identify separate storage and processing time factors from the participant-led and computer-paced administration conditions (i.e. four factors). As the purpose was confirmatory as opposed to exploratory, four factors were forced in the extraction (see Santos et al., 2015 for a similar approach). An orthogonal rotation (Varimax) was used due to the small number of variables in each analysis, as the aim was to create high loadings as opposed to maximising the spread of variables over several factors (Field, 2017).

The Kaiser-Meyer-Olkin (KMO) value was .728 and Bartlett's test of sphericity was significant ($\chi^2(66) = 270.14$, $p < .001$). The range of KMO values for individual items was .50 to .89. The components accounted for 71.84% of the variance. These findings indicated the sample was adequate for PCA. Table 7 shows the results of the PCA for all four variables (participant-led storage and processing time; computer-paced storage and processing time).

Table 7. PCA for storage and processing across conditions

	Factor 1 *E* = 1.55	Factor 2 *E* = 1.71	Factor 3 E= 1.76	Factor 4 *E* = 1.62
Counting span TTC PL	.864			
Counting span PT CP	.859			
Counting span PT PL	776			
Counting span TTC CP	.730			
Odd-one-out PT CP	.520	.509		
Listening span PT CP		.885		
Listening span PT PL		.780		
Odd-one-out span PT PL		.695		
Odd-one-out span TTC PL			.884	
Odd-one-out span TTC CP			.706	
Listening span TTC PL				.899
Listening span TTC CP				.842

PL = Participant-led; CP= computer-paced; TTC = total trials correct (i.e. storage)

The rotation matrix reported in Table 7 shows that the majority of variables loaded on the first two factors. All Counting span variables loaded onto the first factor. The second factor consisted of the processing time variables from the other two tasks, with processing time from Odd-one-out span in the computer-paced condition loading comparably on both the first and second factors. The third and fourth factors contained storage scores for Listening and for Odd-one-out span respectively.

The findings did not show an obvious separation of variables according to whether or not the tasks were computer-paced or participant-led, but suggest that all Counting span measures reflect a single factor. Whilst acknowledging the crossloading for computer-paced Odd-one-out span processing time between the first and second factors, it could be argued that processing time loads to a second factor with the other processing time measures. Components with only two loadings are considered inadequate for use as factors (Field, 2017). However, before discarding storage in the Listening and Odd-one-out span tasks, the same analysis was conducted separately for the participant-led performance and the computer-paced performance indices (see Table 8). This yielded similar results, increasing confidence in the findings.

Table 8. PCA for storage and processing participant-led/computer-paced

	Factor 1	Factor 2	Factor 3	Factor 4
		Participant-led		
	E = 1.55	*E* = 1.71	*E*= 1.76	*E* = 1.62
Counting span TTC	.896			
Counting span PT	.872			
Listening span PT		.916		
Odd-one-out PT		.765		
Odd-one-out TTC			.943	
Listening span TTC				.975
		Computer-paced		
Counting span PT	.871			
Counting span TTC	.776			
Listening span PT		.961		
Odd-one-out PT		.687		
Listening span TTC			.956	
Odd-one-out span TTC				.949

TTC = total trials correct; PL = participant-led; CP = computer-paced; HL = high-level cognition; E = Eigenvalue

The two objectives of this study were: 1) to examine the effect of time-restrictions on CSTs; and 2) to investigate individual contributions of performance indices within CSTs to HLC. The PCAs indicated that Counting span was one factor, and this was the only dimension where the different performance indices share variance, making it possible to address their relative contribution to HLC without introducing bias from variation in associated eigenvalues. Although there was evidence that processing times on the other two tasks form a second factor, and that storage for odd-one-out span and Listening span create two more separate factors, the restricted focus of these on specific indices and tasks rendered them inadequate for use in further analysis. Therefore, it was decided to proceed using the performance indices from Counting span in each condition to represent WM and address the two study objectives.

4. Regression analyses

The first regression analyses examined whether administering the Counting span task in the two administration conditions (participant-led, computer-paced) accounted for the same or different variance in non-verbal reasoning, reading and mathematics. Using a procedure similar to Bailey (2012), separate hierarchical regression analyses were undertaken for each CST performance index (predictor) and each measure of HLC (outcome). Participant-led storage was entered into each regression model at step one, then computer-paced storage was entered at step 2. This analysis was then conducted with computer-paced storage entered at step 1 and participant-led storage at step 2. This indicated the amount of unique variance explained by each variable in each administration condition, when controlling for its counterpart measure. The amount of unique variance for each administration condition was subtracted from the total variance (i.e. the variance explained when scores for both conditions were entered into the model together). The resulting amount of variance was interpreted as the variance shared between the two tasks.

Non-verbal reasoning

Storage in the two administration conditions significantly predicted non-verbal reasoning when they were entered together (F (2,87) = 11.63, p < .001). The amount of total variance accounted for by both scores was 21% (Total R^2 = .21; adjusted = .19, $P < .001$). Looking at the storage scores in each condition separately, computer-paced storage significantly predicted non-verbal reasoning on its own (R^2 = .12, adjusted = .10, p <.01), as did participant-led storage (R^2 = .19, adjusted = .18, p <.001). However, the computer-paced task did not explain variance in non-verbal reasoning when controlling for its counterpart measure (change in R^2 = .02, p = .145). The participant-led task explained variance in non-verbal reasoning above and beyond that explained by the computer-paced task (change in R^2 = .09, p < .01; β = .35, t = 3.13, p < .01). The amounts of variance accounted for by participant-led and computer-paced storage respectively were subtracted from the total variance: i.e. (.21 - .02) - .09 = .10. Variance shared by both storage scores, therefore, was 10%.

Reading

Storage in the two administration conditions significantly predicted reading when entered together (F (2,87) = 4.65, p < .05). The amount of total variance accounted for by both measures was 10% (R^2 = .10, adjusted = .08, p < *.05*). Taking storage in each condition separately, computer-paced storage significantly predicted reading on its own (R^2 = .07, adjusted = .06, p <.01), as did participant-led storage (R^2 = .07, adjusted = .06, p <.05). However, neither accounted for variance in reading when controlling for its counterpart measure (participant-led change in R^2 = .02, p = .17; computer-paced change in R^2 = .03, p = .12). The amounts of variance accounted for by the two measures compared to the total variance was (.10 - .02) - .03 = .05. Variance shared by both measures, therefore, was 5%.

Mathematics

Storage in the two administration conditions significantly predicted mathematics when they were entered together (F (2,87) = 34.71, p < .001). The amount of total variance accounted for by both measures was 44% (R^2 = .44; adjusted = .43. Computer-paced storage significantly predicted mathematics on its own (R^2 = .30, adjusted = .29, p <.001), as did participant-led storage (R^2 = .38; adjusted = .37, p <.001). Computer-paced storage also accounted for variance in mathematics when controlling for its counterpart (change in R^2 = .07, p < .01); as did the participant-led task (change in R^2 = .15, p < .001). Both storage measures held significant relationships with mathematics (participant-led β = .45, t = 4.79, p < .001; computer-paced: β = .31, t = 3.27, p < .01). Variance shared by both measures was (.44 - .07) - .15 = .22; i.e. 22%.

As storage relationships with HLC were not identical in the computer-paced compared with the participant-led conditions of the CST, analysis was conducted next to understand the possible contribution of processing time to variance in measures of HLC above and beyond these. Hierarchical regressions were conducted for each administration condition, with storage entered at Step 1 of the model to control for its contribution to variance in HLC. Then processing time was entered in at Step 2.

Non-verbal reasoning

When processing and storage were put into the model together, they significantly predicted non-verbal reasoning in the participant-led (F (2,87) = 10.85, p < .001; R^2 = .20; adjusted = .18, p < .001) and computer-paced conditions (F (2,88) = 9.39, p < .001; Total R^2 = .18; adjusted = .16, p <.001) conditions. Processing time did not predict non-verbal reasoning above and beyond storage in the participant-led condition (change in R^2 = .01, p = .35). However, in the computer-paced condition, variance was explained by processing time whilst controlling for storage (change in R^2 = .06, p < .05), and processing time was the only variable with a significant relationship with non-verbal reasoning (β = -.28, t = -2.46, p < .05).

Reading

When processing time and storage were put into the model together, they significantly predicted reading in the participant-led (F (2,87) = 3.36, $p < .05$; R^2 = .07; adjusted = .05, $p < .05$) and computer-paced conditions (F (2,88) = 6.31, $p < .01$; Total R^2 = .13; adjusted = .11, $p <.05$) conditions. Processing time did not predict reading above and beyond storage in the participant-led (change in R^2 = .001, p = .72). However, in the computer-paced condition, variance was explained by processing time whilst controlling for storage (change in R^2 = .05, $p < .05$), and processing time was the only variable with a significant relationship with non-verbal reasoning (β = -.29, t = -2.24, $p < .05$).

Mathematics

When the processing time and storage were put into the model together, they significantly predicted mathematics in the participant-led (F (2,87) = 30.78, $p < .001$; R^2 = .41; adjusted = .40, $p < .001$) and computer-paced conditions (F (2,88) = 36.72, $p < .001$; Total R^2 = .46; adjusted = .44, $p <.001$) conditions. Processing time predicted mathematics above and beyond storage in the participant-led condition (change in R^2 = .04, $p < .05$), and both processing time (β = -.26, t = -2.41, $p < .05$) and storage (β = .45, t = 4.15, $p < .001$) showed significant relationships with mathematics. There were similar findings for the computer-paced condition whereby processing time predicted mathematics above and beyond storage (change in R^2 = .16, $p < .001$), and both processing time (β = -.46, t = -5.05, $p < .001$) and storage (β = .32, t = 3.47, $p < .01$) showed significant relationships with mathematics.

These results indicated that, when controlling for storage, additional variance in HLC was explained by processing time within the CST, but, with the exception of mathematics, only in the computer-paced task. This suggests that administration condition is an important factor when considering the contribution of processing time to CST performance.

Discussion

The current study examined 7- and 8-year-old children's WM using CSTs to improve theoretical understanding of the different WM models and subsequent relationships with HLC. There were two objectives: 1) to examine the effects of time-restrictions on CSTs, and 2) to investigate individual contributions of performance indices within CSTs to HLC. The separate CSTs were examined for the effects of time restrictions on individual performance indices compared to the participant-led condition. Then, PCA was conducted to identify factors representing the separate performance indices in order to understand their individual relationships with HLC using hierarchical regression.

Placing time restrictions on the CSTs did not reduce storage scores compared to the tasks with no time restriction (the finding that storage scores in the computer-paced condition of the Listening span task were significantly higher than those in the Participant-led condition will be considered shortly). Given that time restrictions are likely to reduce opportunity for maintenance (Camos & Barrouillet, 2011; Friedman & Miyake, 2004; Lépine et al. 2005; St Clair-Thompson, 2007), this is inconsistent with the multicomponent model (Baddeley & Hitch, 1974), which assumes that WM is reliant on maintenance (e.g. rehearsal), and the TBRS model which assumes reliance on refreshing (Camos & Barrouillet, 2011). Neither does this finding support a fundamental ability limited by attention (Cowan, 1999) as, according to this embedded-process model, storage should increase in the time-restricted condition where there is less interference from individual variation in maintenance strategy use. This is not to say that maintenance is unimportant for encoding information into short-term stores (McNamara & Scott, 2001), but that when there is a concurrent processing task, increased time for maintenance did not improve recall in the current sample of 7- and 8-year-olds.

The absence of impaired storage with processing time restrictions points to the task-switching (Case et al., 1982) and resource-sharing (Towse & Hitch, 1995) accounts of WM, as children were provided with processing time allowance according to their individual speeds. Thus, if resource-sharing explains WM, such a restriction would not impact cognitive resources used for storage. Similarly, if task-switching explains WM, accounting for individual variation in processing speeds would mean time spent away from storage was not increased beyond that required to process the stimuli before switching back to memoranda, thus preventing decay. Furthermore, storage was related to processing times in both conditions for the Counting and Odd-one-out span tasks, consistent with the aforementioned models' supposition that processing speed relates to storage.

Storage and processing times were unrelated in both conditions for Listening span, which may be explained by task-specific stimuli. Unlike Counting and Odd-one-out span, Listening span uses semantic stimuli presented auditorily. Cowan et al. (2003) found that semantic information can be used as a cue in recall, rather than relying solely on phonological memory to recall less meaningful memoranda. This suggests the memoranda are being recalled from long-term memory. As such, a correlation with processing times indicating maintenance, refreshing, resource trade-off or decay prevention would not be expected. It would, however, align to the embedded-process model that posits memoranda in WM are activated from long-term memory. Although the analysis is not included here for the sake of brevity (see mean recall times in Table 2), this explanation is further supported by the considerably longer recall times for Listening span compared to the other two tasks.

This interpretation is also in line with the unexpected finding that mean storage scores for the computer-paced Listening span were significantly higher than mean storage scores for the participant-led version. The stimuli used for the processing components of the participant-led and computer-paced tasks were identical to minimise variation caused by differing processing demands (see St Clair-Thompson, 2007, for a similar methodology with adults). Due to the semantically meaningful nature of sentences, it is possible that some of the sentences were retained in long-term memory from the participant-led trials administered six weeks earlier. Therefore, practice effects may have occurred for this particular task. This could then have resulted in faster processing of the stimuli, thus benefitting time-limited activation, leading to higher span scores (see Cowan et al., 2003, for a similar explanation for longer recall times in a sentence span task).

A further effect of time restrictions was faster processing and recall times, together with poorer processing accuracy, compared to the participant led tasks. The reason for faster processing times – and poorer accuracy – is easily explained by the instruction for children to process the stimuli straight away due to reduced time allowance compared to the participant-led task. The importance of the role of processing speed in HLC is examined below. There are two possible explanations for faster recall times in the computer-paced condition. Participants may have been primed by the faster pace of the processing task, so that they then increased their recall speed. This is feasible, as children may not be able to isolate an instruction to a single component of an overall task (Imeraj et al., 2013). Alternatively, it may be that the computer-paced tasks reduced opportunity for maintenance and encoding into long-term memory (Cowan, 2008). Therefore, with the memoranda still active in short-term memory, participants would attempt to recall the information more rapidly to avoid decay (Cowan & AuBuchon, 2008). In line with this, faster recall times were related to higher storage scores in all three CSTs in the computer-paced condition, but this was only evident for Counting span in the participant-led condition. This is also consistent with the task-switching hypothesis (Towse & Hitch, 1995) which emphasises the role of time-based decay in WM.

These findings suggest that, in relation to the first research question, time for maintenance neither benefits nor disrupts storage in WM, thus supporting either a resource-sharing or a task-switching model. Additionally, a negative linear relationship between storage and processing time provides further evidence for these two models.

The second research question used PCA to examine whether individual performance indices from the CSTs can aid understanding of the WM-HLC relationship, and explain why it is affected by restricted processing times. It is worthy of note that, with one exception (Listening with Odd-one-out span in the participant-led task), correlations between storage scores on the tasks were significant, yet moderate to low. This may indicate that the three tasks tap similar yet independent abilities; suggesting domain-specificity (see Alloway et al., 2003 for a similar explanation).

PCA showed that the Counting span task loaded onto one factor and appeared to best represent WM. Given the small arrays of digits (4-7) in this task, it may be that the processing component relied on subitizing rather than counting (Kaufman, Lord, Reese, & Volkmann, 1949), maximising processing efficiency. Children as young as seven years are well-developed in this skill (Starkey & Cooper, 1995). That Counting span best represented WM is in line with the task-switching model (Towse & Hitch, 1995) as simple processing stimuli are sufficient to draw attention away from storage, thus making the span task complex enough to measure WM. In fact, the TBRS (Barrouillet et al., 2004) and embedded-process (Cowan, 1999) models posit that more complex stimuli bring into play other cognitive abilities that may contaminate the measurement of WM.

Hierarchical regressions demonstrated that the computer-paced and participant-led versions of Counting span measured both similar and different abilities, and this was reflected in relationships with HLC. Reliability analysis indicated that processing accuracy and recall time performance indices from the CSTs were not robust in their representation of single constructs, perhaps because they reflect the unintentional influence of time restrictions, as discussed above. However, processing times had good reliability and explained variance in HLC above and beyond storage in the computer-paced condition.

This finding of the importance of processing speed in the WM-HLC relationship again supports the task-switching model that posits the need to process stimuli quickly in order to prevent decay of memoranda. Given that processing times were faster in the computer-paced condition compared with the participant-led condition, it seems likely that participants with faster processing speeds can only be clearly identified when there is a requirement to process stimuli more quickly, making it possible to isolate the relationship with HLC. This is the first study to provide evidence of this, whilst controlling for individual differences in processing speed. In addition, the reliability of CSTs in two different administration conditions was tested to ensure the same constructs were being measured. No previous study has examined this with children. Also, the current study measured processing speeds within the CSTs, as opposed to using separate tasks, demonstrating that individual differences in processing speed *during* CSTs can explain differences in WM capacity and influence the relationship with HLC.

However, time restrictions weakened relationships with HLC in some instances. For non-verbal reasoning, participant-led storage explained variance beyond that accounted for by computer-paced storage, but not vice versa. Similarly, storage in the participant-led task accounted for twice the amount of variance in mathematics explained by computer-paced storage. These findings suggest that an ability to make use of additional time for maintenance of memoranda is important in HLC, but perhaps because this facilitates downstream comprehension rather than WM storage in itself. This interpretation is contrary to that of Lépine et al. (2005) who argued that maintenance use disrupts the WM-HLC relationship by introducing irrelevant variation in cognitive ability. However, the authors of the current study note that such an interpretation must be applied with caution, as the manipulation of maintenance use is implied, rather than directly measured.

The current study challenges previous research that has found time-restricted CSTs to be better predictors of HLC compared to tasks with no such restriction. This highlights the importance of controlling for individual differences in processing speed when examining WM-HLC relationships. Previous studies finding that time restrictions *strengthen* relationships with HLC have not accounted for individual variation in processing speeds (e.g. Camos & Barrouillet, 2011; Lépine et al. 2005; St Clair-Thompson, 2008). It is possible that generic time restrictions disadvantage children who process stimuli more slowly (i.e. leading to task failure) compared to faster children; and the children who were still able to apply maintenance to the memoranda were those who achieved higher scores on measures of HLC. When that inequality is evened-out by individually titrating the processing time allowance, this (possibly) artefactual relationship is less apparent.

Having ascertained these key points, there would be benefit in extending this study to younger age groups to include those in whom maintenance strategies are less likely to be developed, and in older groups where it is more firmly established. This would enable further understanding of a role (or lack thereof) of maintenance strategy use in the WM-HLC relationship.

Conclusion

The effect of time restrictions on the CSTs provides further evidence for extant theories of WM. An absence of any reduction in storage in time-restricted CSTs challenges models that argue for a role of some form of maintenance in WM (Baddeley & Hitch, 1974; Camos & Barrouillet, 2011; Logie, 1995). The resource-sharing (Case et al., 1982) and task-switching (Towse & Hitch, 1995; Towse et al., 1998) accounts best explain this outcome. Furthermore, findings failed to support the embedded-process model (Cowan, 2005) and TBRS (Lépine et al., 2005) models that posit that time-restricted tasks provide cleaner measures of WM and strengthen links with HLC. Counting span, with its simple processing stimuli, best represented WM, providing further support for the task-switching model and its emphasis on time-based decay rather than resource-sharing. However, participant-led tasks, with slower processing times were better predictors of HLC in some instances. Our interpretation is that faster processing is important to keep information active in WM, in line with the task-switching model (Towse & Hitch, 195); however, explanations of WM that promote factors other than time-based decay are possibly relevant when WM is applied in broader contexts that rely on this resource (e.g. mathematics).

Acknowledgements

The authors would like to thank Sean Rooney from the University's Psychology Laboratory for his skill and dedication in building the tasks reported in this paper. The first author would like to thank Professor Andy Tolmie at Institute of Education, University College London for his invaluable advice on the factor analysis carried out in this paper.

Funding

This work was supported by the Institute for Social Science Research (ISSR) at the relevant university's Centre for Research in Psychology.

References

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133–137.

Andersson, U., & Lyxell, B. (2007). Working memory deficit in children with mathematical difficulties: a general or specific deficit? *Journal of Experimental Child Psychology*, *96*(3), 197–228.

Baddeley, A. D. (1986). *Working Memory.* New York: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley A. D. Hitch G.J. 1974. Working memory. In G. A. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, *Vol. 8* (pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multicomponent model. In 96 MIYAKE ET AL. A. Miyake & P. Shah (Eds.), Models of working memory: Mechanisms of active maintenance and executive control (pp. 28–61). New York: Cambridge Univ. Press.

Bailey, H. (2012). Computer-paced versus experimenter-paced working memory span tasks: Are they equally reliable and valid? *Learning and Individual Differences*, *22*(6), 875–881.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: a time-based resource-sharing model account. *Developmental Psychology*, *45*(2), 477–90.

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*(2), 175.

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology*, *41*(4), 579–97.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71–92.

Berg, D. H. (2008). Working memory and arithmetic calculation in children: the contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology*, *99*(4), 288–308.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273–93.

Camos, V., & Barrouillet, P. (2011). Developmental change in working memory strategies: from passive maintenance to active refreshing. *Developmental Psychology*, *47*(3), 898–904.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786.

Cowan, N. (1999) An embedded-processes model of working memory. In Models of Working Memory (Miyake, A. and Shah, P., eds), pp. 62–101, Cambridge University PressCowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, *24*(1), 154-176.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323-338.

Cowan, N., & AuBuchon, A. M. (2008). Short-term memory loss over time without retroactive stimulus interference. *Psychonomic Bulletin & Review*, *15*(1), 230-235.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42-100.

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science*, *13*(1), 120-133.

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development*, *70*(5), 1082-1097.

Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A. (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science, 18(1),* 1–14.

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., Moreno, M. V., Hitch, G. J. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, *132*(1), 113–132.

Cragg, L., Richardson, S., Hubber, P. J., Keeble, S., & Gilmore, C. (2017). When is working memory important for arithmetic? The impact of strategy and age. *PloS One*, *12*(12), e0188693.

Department for Education (2014) *The national curriculum in England: complete framework for key stages 1 to 4.* Available at: https://www.gov.uk/government/publications/national-curriculum-in-england-framework-for-key-stages-1-to-4 (Accessed: 09 September 2019).

Elliot, C., & Smith, P. (2011). British Ability Scales. *Administration and Scoring Manual* (3rd ed). London: GL assessment.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309-31

Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics: North American Edition*. Sage.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*(1), 136-158.

Gathercole, S. E., Adams, A M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition*, *22*(2), 201–7.

Gathercole, S. E., & Alloway, T. P. (2006). Practitioner review: Short-term and working memory impairments in neurodevelopmental disorders: Diagnosis and remedial support. *Journal of Child Psychology and Psychiatry*, *47*(1), 4-15.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A.M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, 93(3), 265–81.

Gathercole, S. E., Dunning, D. L., & Holmes, J. (2012). Cogmed training: Let's be realistic about intervention research. *Journal of Applied Research in Memory and Cognition,* 1, 201–203.

Gathercole, S. E., & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 189–210). Hove, England: Erlbaum.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *The British Journal of Educational Psychology*, *70(2)*, 177–94.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, *78*(4), 1343–59.

Henry, L. A. (1991). The effects of word length and phonemic similarity in young children's short-term memory. *The Quarterly Journal of Experimental Psychology*, *43*(1), 35-52.

Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory*, *9*(4), 233–247.

Henry, L. (2008). Short-term memory coding in children with intellectual disabilities. *American Journal on Mental Retardation: AJMR*, *113*(3), 187-200.

Henry, L. A., & Botting, N. (2017). Working memory and developmental language impairments. *Child Language Teaching and Therapy*, *33*(1), 19-32.

Henry, L. A, & MacLean, M. (2002). Working memory performance in children with and without intellectual disabilities. *American Journal of Mental Retardation: AJMR*, *107*(6), 421–32.

Henry, L. A., & Millar, S. (1991). Memory span increase with age: A test of two hypotheses. *Journal of Experimental Child Psychology*, *51*(3), 459-484.

Henry, L. A., & Millar, S. (1993). Why does memory span improve with age? A review of the evidence for two current hypotheses. *European Journal of Cognitive Psychology*, *5*(3), 241-287.

Holmes, J., & Gathercole, S. E. (2014). Taking working memory training from the laboratory into schools. *Educational Psychology*, 34(4), 440-450.

Imeraj, L., Antrop, I., Sonuga-Barke, E., Deboutte, D., Deschepper, E., Bal, S., & Roeyers, H. (2013). The impact of instructional context on classroom on-task behavior: A matched comparison of children with ADHD and non-ADHD classmates. *Journal of School Psychology*, *51*(4), 487-498.

Iuculano, T., Moro, R., & Butterworth, B. (2011). Updating Working Memory and arithmetical attainment in school. *Learning and Individual Differences*, *21*(6), 655–661.

Jarrold, C., & Citroën, R. (2013). Reevaluating key evidence for the development of rehearsal: phonological similarity effects in children are subject to proportional scaling artifacts. *Developmental Psychology*, *49*(5), 837.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 498-525.

Kirkup, C., Sizmur, J., Sturman, L., & Lewis, K. (2005). Schools' use of data in teaching and learning (DfES Research Report 671). London: DfES.

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology*, 58(1), 88-111.

Lépine, P. lle, Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review*, *12*(1), 165–170.

Lehmann, M., & Hasselhorn, M. (2007). Variable memory strategy use in children's adaptive intratask learning behavior: Developmental changes and working memory influences in free recall. *Child Development*, *78*(4), 1068-1082.

Logie, R. H. (2014). *Visuo-spatial working memory*. Psychology Press.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, *29*(1), 10-17.

Miller, S., McCulloch, S., & Jarrold, C. (2015). The development of memory maintenance strategies: Training cumulative rehearsal and interactive imagery in children aged between 5 and 9. *Frontiers in Psychology*, *6*, 524.

Passolunghi, M. C., & Cornoldi, C. (2008). Working memory failures in children with arithmetical difficulties. *Child Neuropsychology, 14*(5), 387–400.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, *88*(4), 348–67.

Partanen, P., Jansson, B., Lisspers, J., & Sundin, Ö. (2015). Metacognitive strategy training adds to the effects of working memory training in children with special educational needs. *International Journal of Psychological Studies*, *7*(3), 130-140.

Pickering, S. J., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children*. London: Psychological Corporation.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.

Raven, J. (2008). *Colored Progressive Matrices and Crichton Vocabulary Scale*. London: Pearson Assessment.

Ribner, A. D., Willoughby, M. T., Blair, C. B., & Family Life Project Key Investigators. (2017). Executive function buffers the association between early math and later academic skills. *Frontiers in Psychology*, *8*, 869.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing*, *13*(1-2), 81-103.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Cogmed working memory training: Does the evidence support the claims?. *Journal of Applied Research in Memory and Cognition*, *1*(3), 185-193.

St Clair-Thompson, H. (2007). The influence of strategies on relationships between working memory and cognitive skills. *Memory*, *15*(4), 353–365.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59*(4), 745–59.

Starkey, P., & Cooper Jr, R. G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology*, *13*(4), 399-420.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, *96*(3), 471–491.

Tolmie, A., Muijs, D., & McAteer, E. (2011). *Quantitative methods in educational and social research using SPSS*. McGraw-Hill Education (UK).

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity?. *The Quarterly Journal of Experimental Psychology*, *48*(1), 108-124.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of memory and language*, *39*(2), 195-217.

Towse, J. N., Cowan, N., Horton, N. J., & Whytock, S. (2008). Task experience and children's working memory performance: a perspective from recall timing. *Developmental Psychology*, *44*(3), 695-706.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.